The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1736-1745

SOME PROPERTIES AND AN APPLICATION OF A STATISTIC ARISING IN TESTING CORRELATION[1]

By J. S. Mehta and John Gurland

University of Wisconsin

0. Summary. In testing a hypothesis concerning the correlation coefficient in a bivariate normal distribution where all the parameters are unknown, the Pearson product moment statistic is appropriate. It may happen, however, that there are relations among the parameters in the distribution, in which case the Pearson statistic would not utilize this information. In particular if the variances of the two marginal distributions are equal, it is possible to test the correlation coefficient by means of a simpler statistic which makes use of this information. In this paper we explain how this statistic arises and present some properties of its distribution. This statistic as well as its properties developed here are utilized in the latter part of this paper where we consider the problem of estimating the difference of the means when some of the observations corresponding to one of the variables are missing.

1. Introduction. Let $x_{11}, x_{12}, \cdots, x_{1N}$ and $x_{21}, x_{22}, \cdots, x_{2N}$ be samples from a bivariate normal population having means μ_1, μ_2 and covariance matrix

$$(1) \qquad \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Under the assumption $\sigma_1 = \sigma_2$ the likelihood ratio test of the hypothesis $H_0: \rho = \rho_0$ leads to a critical region based on the statistic u (say) defined as follows

$$(2) \qquad u = 2s_{12}/(s_1^2 + s_2^2)$$

where

$$(3) \qquad Ns_i^2 = \textstyle\sum_{j=1}^N (x_{ij} - \bar{x}_i)^2 \quad \text{for} \quad i = 1, 2,$$

$$(4) \qquad Ns_{12} = \textstyle\sum_{j=1}^N (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2),$$

$$(5) \qquad N\bar{x}_i = \textstyle\sum_{j=1}^N x_{ij}.$$

We present in Section 3 the general non-central distribution of this statistic corresponding to $\sigma_1 \neq \sigma_2$ and $\rho \neq \rho_0$ and we give the first two moments of u in Section 4. These moments are used in Section 5 in investigating the behavior of an estimator of $\mu_1 - \mu_2$ which is proposed here for the case of missing observations corresponding to one of the variables (say) X_2.

Received 15 January 1968; revised 4 April 1969.

[1] This research was supported by NSF Grant GP 7705 and also by the Wisconsin Alumni Research Foundation.

2. A derivation of u. Let us test $H_0: \rho = \rho_0$ against $H_1: \rho \neq \rho_0$. Under the assumption $\sigma_1 = \sigma_2 = \sigma$ (say) the logarithm of the likelihood function can be written down as follows:

$$\begin{aligned}\log L = {} & -N \log 2\pi - N \log \sigma^2 - \tfrac{1}{2}N \log (1 - \rho^2) \\ & - \tfrac{1}{2}N(1 - \rho^2)^{-1}(s_1^2 + s_2^2 - 2\rho s_{12}) - \tfrac{1}{2}N\sigma^2(1 - \rho^2)^{-1} \\ & \cdot[(\bar{x}_1 - \mu_1)^2 + (\bar{x}_2 - \mu_2)^2 - 2\rho(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)].\end{aligned} \tag{6}$$

The solution of the likelihood equations

$$\frac{\partial \log L_1}{\partial \mu_1} = \frac{\partial \log L_1}{\partial \mu_2} = \frac{\partial \log L_1}{\partial \sigma^2} = \frac{\partial \log L_1}{\partial \rho} = 0$$

turns out to be

$$\hat{\mu}_1 = \bar{x}_1; \quad \hat{\mu}_2 = \bar{x}_2; \quad \hat{\sigma}^2 = \tfrac{1}{2}(s_1^2 + s_2^2); \quad \hat{\rho} = 2s_{12}/(s_1^2 + s_2^2). \tag{7}$$

We note here that the maximum likelihood estimate $\hat{\rho}$ of ρ is the statistic u given in (2). The maximum value of the likelihood function under H_1 is

$$\max_{H_1} L(\mu_1, \mu_2, \sigma^2, \rho) = (\pi e)^{-N}[(s_1^2 + s_2^2)^2 - 4s_{12}^2]^{-N/2}. \tag{8}$$

Similarly under the null hypothesis $H_0: \rho = \rho_0$ it can be verified that the maximum likelihood estimates of μ_1, μ_2 and σ^2 are

$$\tilde{\mu}_1 = \bar{x}_1; \quad \tilde{\mu}_2 = \bar{x}_2; \quad \tilde{\sigma}^2 = \tfrac{1}{2}(s_1^2 + s_2^2). \tag{9}$$

The maximum value of the likelihood function under H_0 is

$$\max_{H_0} L(\mu_1, \mu_2, \sigma^2, \rho_0) = (\pi e)^{-N}(1 - \rho_0^2)^{N/2}/(s_1^2 + s_2^2 - 2\rho_0 s_{12})^{-N}. \tag{10}$$

The likelihood criterion is given by

$$\begin{aligned}\lambda = \frac{\max_{H_0} L(\mu_1, \mu_2, \sigma^2, \rho_0)}{\max_{H_1} L(\mu_1, \mu_2, \sigma^2, \rho)} &= \frac{(1 - \rho_0^2)^{N/2}[(s_1^2 + s_2^2)^2 - 4s_{12}^2]^{N/2}}{[s_1^2 + s_2^2 - 2\rho_0 s_{12}]^N} \\ &= [(1 - \rho_0^2)(1 - u^2)(1 - \rho_0 u)^{-2}]^{N/2}\end{aligned} \tag{11}$$

which is a function of u. The likelihood ratio test is

$$(1 - \rho_0^2)(1 - u^2)(1 - \rho_0 u)^{-2} < c \tag{12}$$

where c is chosen so that the probability of the inequality in (12) when samples are drawn from normal populations with correlation ρ_0 is the prescribed significance level α.

The critical region can be written equivalently

$$\begin{aligned} u &> [\rho_0 c + (1 - \rho_0^2)(1 - c)^{\frac{1}{2}}]/(\rho_0^2 c + 1 - \rho_0^2), \\ u &< [\rho_0 c - (1 - \rho_0^2)(1 - c)^{\frac{1}{2}}]/(\rho_0^2 c + 1 - \rho_0^2).\end{aligned} \tag{13}$$

When $\rho_0 = 0$ the critical region becomes

$$(14) \qquad u > (1 - c)^{\frac{1}{2}}, \qquad u < -(1 - c)^{\frac{1}{2}}.$$

3. Distribution of u. The distribution of u both for the null case $\rho = 0$ and the non-null case $\rho \neq 0$ has been given by DeLury (1938) under the assumption $\sigma_1 = \sigma_2$. In this section we derive the non-null distribution without the restriction $\sigma_1 = \sigma_2$.

In order to find the density of u we start from the Wishart density which can be written as

$$(15) \quad p(s_1^2, s_2^2, s_{12}) = K(s_1^2 s_2^2 - s_{12}^2)^{\frac{1}{2}(N-4)} \exp\left[-\tfrac{1}{2}N(1-\rho^2)^{-1}\right] \cdot \left[\frac{s_1^2}{\sigma_1^2} + \frac{s_2^2}{\sigma_2^2} - \frac{2\rho s_{12}}{\sigma_1 \sigma_2}\right]$$

where

$$(16) \qquad K = N^{N-1}[4\pi\Gamma(N-2)\{\sigma_1^2\sigma_2^2(1-\rho^2)\}^{\frac{1}{2}(N-1)}]^{-1}$$

$$0 < s_1^2, s_2^2 < \infty; \qquad -s_1 s_2 < s_{12} < s_1 s_2.$$

To obtain the distribution of u we apply three successive transformations. The first transformation is

$$T_1: u = 2s_{12}/(s_1^2 + s_2^2), \qquad v = s_1^2 + s_2^2, \qquad w = s_2^2.$$

As the joint probability density function of u, v, w is obtained in a straightforward manner, the details are omitted here. Next we apply the transformation

$$T_2: w = \tfrac{1}{2}v[1 + t(1 - u^2)^{\frac{1}{2}}]$$

and integrate out the variable v in the joint density function of u, v, t to obtain the joint probability density function $p(u, t)$ of u and t.

$$(17) \quad p(u,t) \propto \frac{(1-u^2)^{\frac{1}{2}(N-3)}(1-t^2)^{\frac{1}{2}(N-4)}}{\left[\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) + \frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)(1-u^2)^{\frac{1}{2}}t - \frac{\rho u}{\sigma_1 \sigma_2}\right]^{N-1}}$$

$$-1 < u < 1; \; -1 < t < 1.$$

On applying the transformation

$$T_3: t = (1 - z)/(1 + z)$$

we obtain the joint probability density function of u, z which, when z is integrated out yields the probability density function of u

$$(18) \quad p(u) = \frac{(N-2)2^{N-4}(1-\rho^2)^{\frac{1}{2}(N-1)}}{\pi(\sigma_1^2\sigma_2^2)^{\frac{1}{2}(N-1)}}(1-u^2)^{\frac{1}{2}(N-3)} \cdot \left[\frac{2aB(\frac{1}{2}N, \frac{1}{2}(N-2))}{(a^2 - b^2)^{N/2}}\right]$$

which is valid for any sample size $N > 2$. Here

$$(19)\qquad a = \frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - \frac{\rho u}{\sigma_1 \sigma_2}; \qquad b = \frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)(1 - u^2)^{\frac{1}{2}}.$$

When $\sigma_1 = \sigma_2 = \sigma$, say, we have from (19)

$$(20)\qquad a = (1 - \rho u)/\sigma^2; \qquad b = 0.$$

Substitution of these values for a and b in (18) yields

$$(21)\quad p(u) = \pi^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}N)[\Gamma(\tfrac{1}{2}(N-1))]^{-1}(1-\rho^2)^{\frac{1}{2}(N-1)} \cdot (1-u^2)^{\frac{1}{2}(N-3)}(1-\rho u)^{-(N-1)},$$

and when $\rho = 0$ this becomes

$$(22)\qquad p(u) = \pi^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}N)[\Gamma(\tfrac{1}{2}(N-1))]^{-1}(1-u^2)^{\frac{1}{2}(N-3)}.$$

It is evident from (22) that the null distribution of u under these restrictions is the same as that of the product moment correlation coefficient corresponding to $N + 1$ pairs of observations.

4. First two moments of u. It is possible to obtain the moments of u by utilizing directly the distribution of u in Section 3. In this section we obtain these moments through a technique based on moment generating function which yields these results more readily.

Suppose X and Y are random variables with joint moment generating function $\phi_{X,Y}(t_1, t_2)$. If $\Pr\{Y \leqq 0\} = 0$, the mth moment of X/Y is given by

$$(23)\qquad E(X/Y)^m = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \left[\frac{\partial^m}{\partial t_1^m}\phi(t_1, t_2)\right]_{t_1=0} \prod_{j=1}^{m} dt_{2j}$$

where we set $t_2 = \sum_{j=1}^{m} t_{2j}$. (cf. Dixon (1944).)

Now the moment generating function of the Wishart distribution is given by

$$(24)\qquad \phi_{s_1^2, s_2^2, 2s_{12}}(v_{11}, v_{22}, v_{12}) = |I - 2\Sigma V|^{-\frac{1}{2}(N-1)}, \quad \text{where}$$

$$(25)\qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}; \qquad V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}; \qquad \sigma_{12} = \rho\sigma_1\sigma_2 .$$

Consequently the moment generating function $\phi_{X,Y}(t_1, t_2)$ of $X = 2s_{12}$ and $Y = s_1^2 + s_2^2$ is given by

$$\phi_{X,Y}(t_1, t_2) = \phi_{s_1^2, s_1^2, 2s_{22}}(t_2, t_2, t_1)$$

$$(26)\qquad = \left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 2 \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} t_2 & t_1 \\ t_1 & t_2 \end{pmatrix} \right|^{-1/2(N-1)}$$

$$= [1 - 4\sigma_{12}t_1 - 2(\sigma_1^2 + \sigma_2^2)t_2 - 4(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_1^2 + 4(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_2^2]^{-\frac{1}{2}(N-1)}$$

Thus the first two moments of u are given by

$$(27)\quad E(u) = \int_{-\infty}^{0} \left[\frac{\partial\phi(t_1, t_2)}{\partial t_1}\right]_{t_1=0} dt_2$$

$$= \int_{-\infty}^{0} \frac{4(N-1)\sigma_{12}\, dt_2}{[1 - 2(\sigma_1^2 + \sigma_2^2)t_2 + 4(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_2^2]^{\frac{1}{2}(N+1)}}.$$

$$E(u^2) = -\int_{-\infty}^{0} t_2 \left[\frac{\partial^2\phi(t_1, t_2)}{\partial t_1^2}\right]_{t_1=0} dt_2$$

$$(28)\quad = -\int_{-\infty}^{0} \frac{4(N-1)(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_2\, dt_2}{[1 - 2(\sigma_1^2 + \sigma_2^2)t_2 + 4(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_2^2]^{\frac{1}{2}(N+1)}}$$

$$- \int_{-\infty}^{0} \frac{4(N^2-1)\sigma_{12}^2\, t_2\, dt_2}{[1 - 2(\sigma_1^2 + \sigma_2^2)t_2 + 4(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)t_2^2]^{\frac{1}{2}(N+3)}}$$

These integrals can be evaluated by standard techniques utilizing reduction formulae. In Section 5 below we have evaluated these integrals for specific values of N in order to obtain the numerical values of the efficiencies in Table 1.

5. An application of u to incomplete data. In the present section we consider the problem of estimating the difference of the means in a bivariate normal population when the sample has some missing values corresponding to one of the variables. This problem can arise in sample surveys, archeological investigations, psychological tests, and many other situations where it is not always possible or desirable to obtain all of the observations corresponding to both variables (cf. Nicholson (1957)). The problem of estimation of the parameters of a bivariate normal distribution when the sample available is incomplete has been considered by Wilks (1932) and Rao (1952). These authors have obtained the maximum likelihood estimators of the mean vector along with their large sample variances and covariances. More recently, the problem of estimation of the mean vector from an incomplete sample from a trivariate normal population has been considered by Anderson (1957) and Lord (1955). We are concerned here with the problem of estimating the difference of the means. For this, an estimator designated below as Z, in which the u-statistic is utilized, is suggested and the behavior of its variance for small samples is studied.

Suppose that we have a sample

$$(29)\quad \begin{array}{l} x_{11}, x_{12}, \cdots, x_{1n}, x_{1,n+1}, \cdots, x_{1N} \\ x_{21}, x_{22}, \cdots, x_{2n}. \end{array}$$

with $(N - n)$ missing values of the variable X_2 as indicated. Let the underlying bivariate normal population have means μ_1, μ_2 and covariance matrix Σ as in (1). We wish to estimate difference of means δ, say where

$$(30)\quad \delta = \mu_1 - \mu_2.$$

Let us write

$$(31)\quad N\bar{x}_1^{(N)} = \sum_{i=1}^{N} x_{1i}; \qquad n\bar{x}_2^{(n)} = \sum_{i=1}^{n} x_{2i}; \qquad n\bar{x}_1^{(n)} = \sum_{1}^{n} x_{1i};$$
$$(N-n)\bar{x}_1^{(N-n)} = \sum_{n+1}^{N} x_{1i}.$$

A simple estimate of δ which is unbiased is given by

$$(32)\qquad T = \bar{x}_1^{(N)} - \bar{x}_2^{(n)}$$

which has variance

$$(33)\qquad \text{Var}\,(T) = n^{-1}\sigma_1^2[\lambda + k - 2\rho\lambda(k)^{\frac{1}{2}}]$$

where

$$(34)\qquad k = \sigma_2^2/\sigma_1^2; \qquad \lambda = n/N.$$

There are, however, other unbiased estimators of δ which, although not as simple as T, have variance less than that of T for certain values of the parameters involved, namely ρ, k and λ.

One such estimator of δ can be obtained as follows. Consider the estimator Z_0 defined by

$$(35)\qquad Z_0 = A_1\bar{x}_1^{(n)} + A_2\bar{x}_1^{(N-n)} - \bar{x}_2^{(n)}$$

which will be unbiased for δ if $A_2 = 1 - A_1$. Accordingly we write

$$(36)\qquad Z_0 = A_1\bar{x}_1^{(n)} + (1 - A_1)\bar{x}_1^{(N-n)} - \bar{x}_2^{(n)}$$

where A_1 is an arbitrary constant to be specified. A plausible method of specifying A_1 is to choose that value which minimizes the variance of Z_0. It is easily shown that the variance of Z_0 is

$$(37)\qquad \text{Var}\, Z_0 = n^{-1}\sigma_1^2[A_1^2 + (1 - A_1)^2\lambda(1-\lambda)^{-1} + k - 2\rho A_1(k)^{\frac{1}{2}}]$$

and the value of A_1 which minimizes (37) is given by

$$(38)\qquad A_1 = [n + \rho(k)^{\frac{1}{2}}(N - n)]/N = \lambda + \rho(k)^{\frac{1}{2}}(1 - \lambda).$$

Of the three parameters λ, ρ and k involved in (38), λ will generally be the only one known in any given practical situation. Thus to employ Z_0 in practice as an estimator of δ we must specify ρ and k.

Now, as already shown, u is the maximum likelihood estimate of ρ when $k = 1$. Consequently one would expect that the estimator Z defined as

$$(39)\qquad Z = A\bar{x}_1^{(n)} + (1 - A)\bar{x}_1^{(N-n)} - \bar{x}_2^{(n)}$$

with

$$(40)\qquad A = \lambda + u(1 - \lambda)$$

would perform well in the neighborhood of $k = 1$. As will become evident in the following sections it turns out that this estimator has a smaller variance than the estimator T, not only in the neighborhood of $k = 1$ but also in certain other neighborhoods, depending on the value of the correlation coefficient ρ and the number of missing values.

6. Efficiency of Z. Since Z is an unbiased estimator of $\mu_1 - \mu_2$ a natural way of evaluating its behavior is to compare its variance with that of T. For this purpose we define the efficiency of Z as

$$\text{eff } Z = \text{Var } T/\text{Var } Z. \tag{41}$$

We obtain the variance of Z as follows, utilizing the fact that Z is conditionally unbiased given u.

$$\begin{aligned} \text{Var } Z &= E_u[\text{Var }(Z \mid u)] \\ &= \sigma_1^2 n^{-1}(1-\lambda)^{-1} \\ &\quad \cdot[\lambda + (1-\lambda)k - \{2\lambda + (1-\lambda)\rho(k)^{\frac{1}{2}}\}E(A) + E(A^2)] \end{aligned} \tag{42}$$

where

$$\begin{aligned} E(A) &= \lambda + (1-\lambda)E(u) \\ E(A^2) &= \lambda^2 + 2\lambda(1-\lambda)E(u) + (1-\lambda)^2E(u^2). \end{aligned} \tag{43}$$

Consequently the efficiency of Z is

$$\text{eff } Z = \frac{(\lambda + k - 2\rho(k)^{\frac{1}{2}}\lambda)(1-\lambda)}{[\lambda + (1-\lambda)k - 2\{\lambda + (1-\lambda)\rho k^{\frac{1}{2}}\}E(A) + E(A^2)]}. \tag{44}$$

It can be easily shown that eff $Z \gtrless 1$ according as

$$E(u^2) \lessgtr 2\rho(k)^{\frac{1}{2}}E(u). \tag{45}$$

Since $E(u^2) > 0$, there is a loss in efficiency if $\rho = 0$. By the continuity of the expression for efficiency there is a small interval around $\rho = 0$ for which Z will continue to be inefficient. The behavior of Z when ρ is not in the neighborhood of zero is not so obvious and will be examined in some detail below. At the same time the dependence of efficiency on values of k will also be examined.

We have calculated the efficiency of the estimator Z for several values of n and also for a grid of values of the parameters ρ, λ and k. As the results for moderate values of n are all quite similar it will suffice to present here the values of the efficiency for $n = 17$ only. In Table 1 appear the values of the efficiency for the indicated values of ρ, λ and k.

On examining the table it is evident that there is a gain in efficiency in some regions of the parameter space and a loss in others. There appears to be a general loss of efficiency in the region around $\rho = 0$ extending from about $\rho = -.1$ to $\rho = 0.1$ with a few exceptions where there is a very slight gain. In the remaining region of ρ extending beyond $|\rho| > 0.1$ there appears to be a general gain in efficiency except for a few instances where there is a slight loss.

For fixed values of k and ρ the gain in efficiency decreases as λ increases from

TABLE 1

Efficiency of Z when n = 17

λ	ρ	k: 0.1	0.2	0.5	1.0	2.0	5.0	10.0
0.1	−.9	1.0942	1.2645	1.8046	2.3104	2.2014	1.5752	1.2898
	−.7	1.0306	1.1231	1.3595	1.5182	1.4899	1.2828	1.1574
	−.5	.9773	1.0181	1.1225	1.1897	1.1918	1.1246	1.0741
	−.3	.9382	.9483	.9941	1.0313	1.0474	1.0379	1.0244
	−.1	.9151	.9110	.9346	.9629	.9854	.9986	1.0010
	.1	.9101	.9061	.9316	.9616	.9850	.9986	1.0010
	.3	.9262	.9388	.9932	1.0351	1.0516	1.0400	1.0254
	.5	.9690	1.0245	1.1604	1.2366	1.2258	1.1376	1.0793
	.7	1.0488	1.2011	1.5846	1.7885	1.6591	1.3321	1.1744
	.9	1.1828	1.5689	3.1843	4.7419	3.2928	1.7475	1.3361
0.5	−.9	1.0141	1.0418	1.1237	1.2020	1.2334	1.1745	1.1109
	−.7	1.0049	1.0223	1.0738	1.1197	1.1367	1.1036	1.0668
	−.5	.9961	1.0037	1.0310	1.0561	1.0667	1.0521	1.0342
	−.3	.9886	.9883	0.9982	1.0110	1.0191	1.0175	1.0120
	−.1	.9835	.9780	.9782	.9852	.9934	.9993	1.0005
	.1	.9816	.9751	.9749	.9832	.9926	.9993	1.0005
	.3	.9844	.9828	.9973	1.0166	1.0272	1.0225	1.0145
	.5	.9933	1.0072	1.0671	1.1189	1.1258	1.0808	1.0469
	.7	1.0107	1.0605	1.2552	1.4163	1.3850	1.2026	1.1064
	.9	1.0405	1.1746	1.9820	3.0507	2.3907	1.4708	1.2107
0.9	−.9	1.0016	1.0049	1.0144	1.0235	1.0283	1.0240	1.0169
	−.7	1.0006	1.0027	1.0090	1.0151	1.0183	1.0154	1.0108
	−.5	.9995	1.0005	1.0040	1.0076	1.0097	1.0084	1.0058
	−.3	.9986	.9985	.9998	1.0016	1.0030	1.0030	1.0022
	−.1	.9980	.9972	.9969	.9977	.9989	.9999	1.0001
	.1	.9978	.9967	.9963	.9972	.9987	.9999	1.0001
	.3	.9981	.9977	0.9996	1.0029	1.0052	1.0045	1.0030
	.5	.9992	1.0010	1.0108	1.0217	1.0253	1.0171	1.0100
	.7	1.0001	1.0083	1.0420	1.0793	1.0812	1.0450	1.0236
	.9	1.0051	1.0241	1.1649	1.4047	1.3063	1.1087	1.0483

0.1 to 0.9. On the other hand the loss in efficiency decreases as λ increases. It is also apparent that whatever be the values of k and λ the efficiency increases as ρ moves away from zero in negative or positive direction. Further, for values of ρ close to ± 1 and whatever be the value of λ the efficiency appears to attain a maximum for values of k in some interval around $k = 1$. This is not surprising, however, since k was assigned the value one in constructing an estimator Z with minimum variance.

In conclusion, it is evident that the estimator Z has a substantially smaller variance than that of T if the values of the parameter ρ are sufficiently far away from zero. It was, in fact, for this purpose that the estimator z was constructed, namely, to take advantage of the possible correlation between the two samples when such correlation may exist. Although in the construction of this statistic Z

the ratio k of variances was taken to be one, it turns out that the variance of Z is still substantially less than that of T in a wide range of k values when the correlation is sufficiently large numerically. Consequently the statistic u is not only interesting in itself as a simpler estimator than the product moment correlation coefficient, but it can also be utilized advantageously in the problem of estimating a difference of means.

7. Use of Z in testing $H_0:\delta = \delta_0$. In Section 5 we considered the statistic

$$Z = A\bar{x}_1^{(n)} + (1 - A)\bar{x}_1^{(N-n)} - \bar{x}_2^{(n)}$$

with variance

$$V(Z) = n^{-1}(1 - \lambda)^{-1}[\lambda\sigma_1^2 + (1 - \lambda)\sigma_2^2 - 2\{\lambda\sigma_1^2 + (1 - \lambda)\rho\sigma_1\sigma_2\}E(A) + \sigma_1^2E(A^2)].$$

It should be noted that the parameters involved in $V(Z)$ are $\sigma_1^2, \sigma_2^2, \sigma_{12}(=\rho\sigma_1\sigma_2)$. Now under H_0, $[V(Z)]^{-\frac{1}{2}}(Z - \delta)$ has a standardized normal distribution $N(0, 1)$. If $\sigma_1^2, \sigma_2^2, \sigma_{12}$ are unknown one can use the modified statistic

$$Z^* = [\hat{V}(Z)]^{-\frac{1}{2}}(Z - \delta) \tag{46}$$

where now s_1^2, s_2^2, s_{12}, the maximum likelihood estimators of $\sigma_1^2, \sigma_2^2, \sigma_{12}$ respectively have been substituted in the expression for $V(Z)$ to give $\hat{V}(Z)$. Since the maximum likelihood estimators are consistent, therefore, as n, and consequently $N - n$, tend to infinity the distribution of Z^* tends to $N(0, 1)$. The referee has suggested the possibility of utilizing Z^* for testing H_0, and this aspect is currently under investigation.

8. Example. Now we illustrate the use of the estimator Z through an example. A sample of size 10 was drawn from a bivariate normal population with mean vector $\left[\begin{smallmatrix}4\\6\end{smallmatrix}\right]$ and covariance matrix $\left[\begin{smallmatrix}1 & 0.8\\0.8 & 4\end{smallmatrix}\right]$. Let us assume that the last five observations on X_2 are missing. The quantities required in calculating Z are

$$\bar{x}_1^{(n)} = 4.956; \quad \bar{x}_1^{(N-n)} = 3.160; \quad \bar{x}_2^{(n)} = 6.238;$$

$$\lambda = 0.5; \quad \textstyle\sum_1^n (x_{1i} - \bar{x}_1^{(n)})^2 = 4.773;$$

$$\textstyle\sum_1^n (x_{2i} - \bar{x}_2^{(n)})^2 = 25.266; \quad \sum_1^n (x_{1i} - \bar{x}_1^{(n)})(x_{2i} - \bar{x}_2^{(n)}) = 4.484.$$

This yields $u = 0.299$, $A = 0.649$ and consequently $Z = -1.912$.

For the purpose of comparison we note that $T = \bar{x}_1^{(n)} - \bar{x}_2^{(n)} = -2.180$. Evidently for this sample the estimate obtained by using Z is closer to the true value $\mu_1 - \mu_2 = -2.0$ than the simple estimate obtained by using T.

For further comparison we consider the estimate which would have been obtained by taking the difference $\bar{x}_1^{(10)} - \bar{x}_2^{(10)}$ if all 10 observations on both variables had been available. From the sample we obtain $\bar{x}_1^{(10)} - \bar{x}_2^{(10)} = -2.049$. This is closer to the true difference $\mu_1 - \mu_2 = -2.0$. However the performance of Z in the absence of 5 observations indicates some advantage over the use of the simple difference of the means when such data are missing.

REFERENCES

[1] Afifi, A. and Elashoff, R. M. (1966). Missing observations in multivariate statistics I. Review of literature. *J. Amer. Statist. Assoc.* **61** 595–604.

[2] Anderson, R. L. (1942). Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13** 1–13.

[3] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[4] Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52** 200–204.

[5] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[6] DeLury, D. B. (1938). Note on correlations. *Ann. Math. Statist.* **9** 149–51.

[7] Dixon, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Statist.* **15** 119–44.

[8] Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series and Products*. Academic Press, New York.

[9] Lord, F. M. (1955). Estimation of parameters from incomplete data. *J. Amer. Statist Assoc.* **50** 870–876.

[10] Morgan, W. A. (1939). A test for significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika* **31** 13–19.

[11] Nicholson, G. E. (1957). Estimation of parameters from incomplete multivariate samples. *J. Amer. Statist. Assoc.* **52** 523–526.

[12] Pitman, E. J. (1939). A note on normal correlation. *Biometrika* **31** 9–12.

[13] Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. Wiley, New York. 161–165.

[14] Wilks, S. S. (1932). Moments and distribution of estimates of population parameters from fragmentary samples. *Ann. Math. Statist.* **3** 163–203.